Linking secondary metabolites to gene clusters through genome sequencing of six diverse Aspergillus species

Kjærbølling, Inge; Vesth, Tammi C.; Frisvad, Jens C.; Nybo, Jane L.; Theobald, Sebastian; Kuo, Alan; Bowyer, Paul; Matsuda, Yudai; Mondo, Stephen; Lyhne, Ellen K.

Total number of authors:
27

Published in:
Proceedings of the National Academy of Sciences of the United States of America

Link to article, DOI:
10.1073/pnas.1715954115

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Linking secondary metabolites to gene clusters through genome sequencing of six diverse *Aspergillus* species

Inge Kjærbølling[a], Tammi C. Vesth[a], Jens C. Frisvad[a], Jane L. Nybo[a], Sebastian Theobald[a], Alan Kuo[b], Paul Bowyer[c], Yudai Matsuda[a], Stephen Mondo[b], Ellen K. Lyhne[a], Martin E. Kogle[a], Alicia Clum[b], Anna Lipzen[b], Asaf Salamov[b], Chew Yee Ngan[b], Chris Daum[b], Jennifer Chiniquy[b], Kerrie Barry[b], Kurt LaButti[b], Sajeet Haridas[b], Blake A. Simmons[d,e], Jon K. Magnuson[d,f], Uffe H. Mortensen[a], Thomas O. Larsen[a], Igor V. Grigoriev[b,g], Scott E. Baker[d,h], and Mikael R. Andersen[a,1]

[a]Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Lyngby, Denmark; [b]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598; [c]Manchester Fungal Infection Group, Institute of Inflammation and Repair, Faculty of Medicine and Human Sciences, University of Manchester, Manchester M13 9PL, United Kingdom; [d]US Department of Energy Joint BioEnergy Institute, Emeryville, CA 94608; [e]Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [f]Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA 99352; [g]Plant and Microbial Biology Department, University of California Berkeley, Berkeley, CA 94720; and [h]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved December 8, 2017 (received for review September 11, 2017)

The fungal genus of *Aspergillus* is highly interesting, containing everything from industrial cell factories, model organisms, and human pathogens. In particular, this group has a prolific production of bioactive secondary metabolites (SMs). In this work, four diverse *Aspergillus* species (*A. campestris*, *A. novofumigatus*, *A. ochraceoroseus*, and *A. steynii*) have been whole-genome PacBio sequenced to provide genetic references in three *Aspergillus* sections. *A. taichungensis* and *A. candidus* also were sequenced for SM elucidation. Thirteen *Aspergillus* genomes were analyzed with comparative genomics to determine phylogeny and genetic diversity, showing that each presented genome contains 15–27% genes not found in other sequenced Aspergilli. In particular, *A. novofumigatus* was compared with the pathogenic species *A. fumigatus*. This suggests that *A. novofumigatus* can produce most of the same allergens, virulence, and pathogenicity factors as *A. fumigatus*, suggesting that *A. novofumigatus* could be as pathogenic as *A. fumigatus*. Furthermore, SMs were linked to gene clusters based on biological and chemical knowledge and analysis, genome sequences, and predictive algorithms. We thus identify putative SM clusters for aflatoxin, chlorflavonin, and ochrindol in *A. ochraceoroseus*, *A. campestris*, and *A. steynii*, respectively, and novofumigatonin, *ent*-cycloechinulin, and *epi*-aszonalenins in *A. novofumigatus*. Our study delivers six fungal genomes, showing the large diversity found in the *Aspergillus* genus; highlights the potential for discovery of beneficial or harmful SMs; and supports reports of *A. novofumigatus* pathogenicity. It also shows how biological, biochemical, and genomic information can be combined to identify genes involved in the biosynthesis of specific SMs.

Aspergillus | fumigatus | comparative genomics | secondary metabolism

The *Aspergillus* genus is a diverse group of fungal species found worldwide in varying habitats. Several species are used in biotechnological industries for the production of enzymes and metabolites (commodity chemicals and pharmaceuticals), and as fermentation agents in food (1). Certain species, such as *A. clavatus* and *A. fumigatus*, are known food spoilers, mycotoxin producers, and opportunistic pathogens (1, 2). To study this diversity, it is important to have reference genomes of high assembly quality in all major clades of the genus. For this purpose, we selected four diverse *Aspergillus* species, *A. campestris*, *A. novofumigatus*, *A. ochraceoroseus*, and *A. steynii*, representing four phylogenetically very different sections in *Aspergillus*, for high-quality PacBio sequencing. The four selected genomes represent diverse and genomically unexplored sections of the *Aspergillus* genus: *A. campestris* is the first member of section *Candidi* to be sequenced, and likewise *A. steynii* is the first member of section *Circumdati* to be sequenced. *A. ochraceoroseus*, the first member of section *Ochraceorosei*, has recently been draft genome sequenced (3) and is available only in a large number of scaffolds. Here we also present a greatly improved assembly that may serve as a reference genome for this section. Furthermore, we have added a highly interesting member of section *Fumigati*, *A. novofumigatus*, which has a diverse secondary metabolite (SM) profile (4), as well as potentially being an opportunistic pathogen with close relation to the medically very important *A. fumigatus* (5). In addition, two strains from the *Candidi* section were Illumina sequenced to elucidate the chlorflavonin biosynthesis.

Significance

The genus of *Aspergillus* holds fungi relevant to plant and human pathology, food biotechnology, enzyme production, model organisms, and a selection of extremophiles. Here we present six whole-genome sequences that represent unexplored branches of the *Aspergillus* genus. The comparison of these genomes with previous genomes, coupled with extensive chemical analysis, has allowed us to identify genes for toxins, antibiotics, and anticancer compounds, as well as show that *Aspergillus novofumigatus* is potentially as pathogenic as *Aspergillus fumigatus*, and has an even more diverse set of secreted bioactive compounds. The findings are of interest to industrial biotechnology and basic research, as well as medical and clinical research.

Author contributions: I.K., T.C.V., J.C.F., E.K.L., M.E.K., K.B., B.A.S., J.K.M., U.H.M., T.O.L., I.V.G., S.E.B., and M.R.A. designed research; I.K., T.C.V., E.K.L., M.E.K., C.Y.N., C.D., and J.C. performed research; I.K., A.S., B.A.S., J.K.M., and T.O.L. contributed new reagents/analytic tools; I.K., T.C.V., J.C.F., J.L.N., S.T., A.K., P.B., Y.M., S.M., A.C., A.L., A.S., K.L., S.H., T.O.L., S.E.B., and M.R.A. analyzed data; and I.K., T.C.V., J.C.F., J.L.N., S.T., P.B., Y.M., S.M., M.E.K., U.H.M., T.O.L., I.V.G., and M.R.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All the sequencing data are available at the JGI Genome Portal (genome.jgi.doe.gov). *A. campestris* (accession no. MSFM00000000) genome.jgi.doe.gov/Aspcam1/Aspcam1.home.html *A. novofumigatus* (accession no. MSZS00000000) genome.jgi.doe.gov/Aspnov1/Aspnov1.home.html *A. ochraceoroseus* (accession no. MSFN00000000) genome.jgi.doe.gov/Aspoch1/Aspoch1.home.html *A. steynii* (accession no. MSFO00000000) genome.jgi.doe.gov/Aspste1/Aspste1.home.html A. candidus (accession no. PKFS00000000) genome.jgi.doe.gov/Aspcand1/Aspcand1.home.html *A. taichungensis* (accession no. PKFW00000000) genome.jgi.doe.gov/Asptaic1/Asptaic1.home.html.

[1]To whom correspondence should be addressed. Email: mr@bio.dtu.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1715954115/-/DCSupplemental.

The four PacBio sequenced species explored in this study can act as a reference strain in their respective phylogenetic sections. The species may also be used to assess the natural variation within the *Aspergillus* genus via analysis of species-specific genes in comparison with other genome-sequenced species. Accordingly, we have compared our sequenced genomes with nine published *Aspergillus* reference genomes (from sections *Nidulantes*, *Nigri*, *Fumigati*, *Flavi*, *Clavati*, and *Terrei*) to serve as a compilation of reference strains for the genus.

In addition, the biosynthetic potential of these species is of interest. Filamentous fungi produce a diverse range of SMs, including bioactive compounds such as pharmaceuticals and toxins (6). SMs are not required for growth, but provide important benefits in the growth environment (7). Members of the *Aspergillus* genus are known to produce a wide variety of SMs with industrial, agricultural, medical, and economic importance (7, 8). The biosynthetic genes of SMs are located in clusters setting the stage for common gene regulation (9, 10). Clusters often span tens of kilobases (kbs) (11) and usually contain a gene or genes coding for one or more synthases (backbone enzyme) that define the product class of the cluster [i.e., polyketide synthases (PKS), nonribosomal peptide synthetases, and prenyltransferases or terpene cyclases (12)], in addition to tailoring enzymes such as transferases, hydroxylases, and regulatory proteins and transporters (11, 12).

With the increasing number of whole-genome sequences, the opportunity of performing analysis based on comparative genomics arises, which can give important insights and knowledge. With a focus on investigating bioactive and toxic compounds, we have here identified biosynthetic gene clusters responsible for interesting compounds from each of the PacBio sequences by combining genome analysis with knowledge of biochemical pathways and compound structure. We have identified candidates for the ochrindol cluster in *A. steynii*, and the chlorflavonin cluster in *A. campestris*.

A. novofumigatus was investigated on a genetic level, focusing on SMs. The secondary metabolic potential has been investigated, and biosynthetic gene clusters for three compounds (novofumigatonin, *epi*-aszonalenin, and *ent*-cycloechi) have been identified. Furthermore, the genomic differences and similarities of the closely related species *A. novofumigatus* and the pathogen *A. fumigatus* have been investigated, focusing on SMs, allergens, and virulence factors, and thereby addressing the potential pathogenicity of *A. novofumigatus* and how closely related the two morphologically similar species are.

In addition, the evolution of the aflatoxin (a highly carcinogenic compound) gene cluster from *A. ochraceoroseus* was investigated. The biosynthetic gene cluster was identified and studied earlier in several species, including *A. flavus*, *A. parasiticus*, and *A. ochraceoroseus* (13, 14). It has been seen that the synteny of the clusters are quite varying and that *A. ochraceoroseus* is missing some essential genes (*aflQ* and *aflP*) in the biosynthesis of aflatoxin known from *A. flavus* (14). With whole-genome sequences at hand, we have addressed some of these questions concerning the evolution of this biosynthetic gene cluster.

Results and Discussion

Genome Statistics. The genomes of *A. campestris*, *A. novofumigatus*, *A. ochraceoroseus*, and *A. steynii* were sequenced using PacBio RS, whereas *A. taichungensis* and *A. candidus* were sequenced using Illumina (see *SI Appendix* for details). Annotation of the genomes was completed using the JGI Annotation Pipeline (15). Table 1 lists genome sequence statistics for each of the six species. The four PacBio sequenced genomes have a relatively low number of scaffolds and do not contain internal gaps. For that reason, they are highly useful as references for comparative genomics, as well as for studies of the individual genomes. Of the sequenced genomes, *A. steynii* has the largest genome size and is comparable with that of *A. oryzae* (16). The genome of *A. steynii* is ~27% larger than *A. ochraceoroseus*, which has the smallest genome in this set and has a genome size comparable with *A. clavatus* (17). The difference in genome size also reflects the numbers of predicted genes in the two species, which range from 13,211 to 8,924, respectively.

Investigation of DNA Methylation. Because the four *Aspergillus* genomes (*A. steynii*, *A. campestris*, *A. novofumigatus*, *A. ochraceoroseus*) have been sequenced using PacBio, it is possible to investigate the presence of N6-methyldeoxyadenine (6mA) (18). Previous attempts at validation of such low abundance of 6mA have proven challenging, making it difficult to conclude whether 6mA is present in these fungi and, if so, to discriminate between real 6mA sites and false-positives (18). The presence of 6mA was therefore explored across the four *Aspergillus* genomes (Table 2). Consistent with previous reports (18) of low levels of 6mA in the Dikarya, we detect very little 6mA in the Aspergilli, ranging from 0.012 (*A. steynii*) to 0.038 (*A. campestris*) percent adenines methylated compared with early-diverging fungi, in which up to 2.8% of all adenines were methylated (Table 2) (18). Furthermore, only a handful of 6mA sites were at ApT dinucleotides, and none was found symmetrically at ApTs, both of which are characteristic features of 6mA modification in early-diverging fungi (18). The results therefore suggest an absence or very low occurrence of 6mA methylation in Aspergilli.

Whole-Genome Phylogeny Confirms Species Found in Separate Clades. To provide an overview of the relationships among the sequenced species in the *Aspergillus* genus, we constructed a phylogenetic tree of the four PacBio sequenced species and the 11 reference strains, including *Penicillium chrysogenum* and *Neurospora crassa* as outgroups (Fig. 1).

The constructed phylogenetic tree supports the results described earlier by Peterson (21), where a tree was constructed based on DNA sequences of four loci. *A. campestris* most closely resembles *A. terreus* of the reference genomes, whereas *A. steynii* relates closest to *A. flavus* and *A. oryzae*. Members of the *Fumigati*

Table 1. Overview of sequencing and annotation data for the four investigated PacBio-sequenced species, plus two additional Illumina-sequenced species

	A. campestris	*A. novofumigatus*	*A. ochraceoroseus*	*A. steynii*	*A. candidus*	*A. taichungensis*
Genome size, Mbp	28.3	32.4	27.7	37.8	27.3	27.12
Number of proteins	9,764	11,549	8,924	13,211	9,641	9,692
Number of scaffolds	62	62	34	37	268	310
Number of scaffolds ≥2 kbp	56	62	32	36	168	283
Scaffold N50	6	4	4	4	23	47
Scaffold L50	1,703,432	3,768,347	2,489,623	3,921,250	391,998	207,690
Fraction of GC, %	51.2	49.1	44.2	49.1	51.8	51.44
Coverage of gaps, %	0	0	0	0	0.0298	0.0155
Coverage of InterPro, %	68	67	67	66	75	25

Table 2. Overview of the methylation pattern of *A. campestris*, *A. novofumigatus*, *A. ochraceoroseus*, and *A. steynii*

Lineage	Percentage adenines methylated	Total number of sites	Percentage modifications at ApT sites
A. steynii	0.012	6,753	0.054
A. campestris	0.038	9,156	0.041
A. novofumigatus	0.03	7,917	0.058
A. ochraceoroseus	0.021	7,355	0.027

section are in a single clade (marked in blue on Fig. 1), with *A. clavatus* as a close relative. *A. ochraceoroseus* is placed next to *A. nidulans*, and both belong to the subgenus Nidulantes. All the species belonging to subgenus Circumdati (*A. niger*, *A. oryzae*, *A. flavus*, *A. steynii*, *A. terreus*, and *A. campestris*) are also placed in one clade. The tree further confirms that the three species *A. ochraceoroseus*, *A. steynii*, and *A. campestris* indeed represent distinct branches in the *Aspergillus* phylogram (22).

Unique Genes in the Genomes Often Encode Regulatory Proteins and Enzymes Involved in Secondary Metabolism. We have identified and investigated species-specific genes for the four newly sequenced species to examine the diversity within the *Aspergillus* genus. Genes that are unique to a species or a small group of species may be associated with phenotypic traits and adaptation of these species to specific environments. We define species-specific genes as those without any orthologs in other sequenced genomes. This definition makes the set of species-specific genes dependent on the strains included in the analysis. As more genomes are included, especially genomes from closely related species or strains, fewer species-specific genes will be identified. The species-specific genes for each genome were identified using a set consisting of the four PacBio sequenced genomes and 11 reference genomes (*SI Appendix*, Table S1). Two closely related strains will share most of their genes, and they will as such not be unique to the individual species. The unique genes are not expected to encode any key functions in the cell, as they are found in only one organism; instead, these genes might be involved in environmental adaptation and/or speciation. The strains have 22%, 15%, 21%, and 27% unique genes for *A. campestris*, *A. novofumigatus*, *A. ochraceoroseus*, and *A. steynii*, respectively, indicating the vast diversity found within the *Aspergillus* genus. Approximately one third of the species-specific genes could be associated with an InterPro sequence domain (*SI Appendix*, Table S1) (23), suggesting that these genes are not false annotations.

Comparative Analysis of the Genomes of *A. novofumigatus* and *A. fumigatus*. *A. novofumigatus* and *A. fumigatus* are considered to be two closely related species, and *A. novofumigatus* has only been regarded a separate species since 2005 (24). The homology between *A. novofumigatus* and *A. fumigatus* has been investigated based on the number of *A. novofumigatus* proteins with BLASTP hits (≥50% identity ≥130% coverage of query plus hit) in *A. fumigatus*. Based on this, 8,385 of *A. novofumigatus* proteins have homologous proteins in *A. fumigatus*, corresponding to 73%. The synteny between the two species was also examined using NUCmer (Nucleotide Mummer) from the MUMmer 3.0 package to map *A. novofumigtus* genome to the reference genome of *A. fumigatus* (25–27). Based on these alignments, 23.1 Mbp of the *A. novofumigatus* genome can be mapped to *A. fumigatus*, corresponding to 71% of the *A. novofumigatus* genome. The maximum block size is 75 kbp, and the mean block size is 4.6 kbp.

To explore this difference genetically and functionally, we have explored the similarities and differences between these two species with a focus on allergens, genes involved in virulence, and production of SMs.

Secondary Metabolite Profile of *A. novofumigatus* Compared with *A. fumigatus*. The extrolite production in *A. fumigatus* has been extensively studied, and an abundance of SMs have been identified (4). *A. novofumigatus* is also known to have a versatile secondary metabolism; however, there is very little overlap of extrolite production between the two closely related species, making it very interesting to compare their genetic potential for producing SMs (4).

To investigate what type of SM gene clusters *A. fumigatus* and *A. novofumigatus* have in common, the SM gene clusters were predicted for each genome, using an implementation of SMURF (28). An overview of the predicted clusters and homologs in *A. novofumigatus* and *A. fumigatus* is presented in Fig. 2 *A* and *B*, respectively. Of the 34 predicted clusters in *A. fumigatus* and 56 predicted clusters in *A. novofumigatus*, 24 appear to be shared among the two species, based on bidirectional BLAST hits of the synthase (Fig. 2*C*). Of the 11 elucidated clusters from *A. fumigatus* [based on MIBiG (29)], homologs of seven (Gliotoxin, hexadehydro-astechrome, pseurotin A, fumagillin, endocrocin, helvolic acid, and trypacidin) can be found in *A. novofumigatus* (based on homology of the synthase). Several of these SMs are known to be involved in the virulence of *A. fumigatus* and are examined in more detail in the *Comparative Genomics of Genes Encoding Allergens, Virulence, and Pathogenicity Factors*.

The prediction of SM gene clusters also revealed considerable differences between the two closely related species. First, as seen in Fig. 2*A*, *A. novofumigatus* has 17 predicted clusters with no orthologs in any of the reference species. This is in contrast to *A. fumigatus*, which has only three clusters without orthologs in the reference species (Fig. 2*B*). Second, *A. novofumigatus* has 56 predicted SM clusters, whereas *A. fumigatus* has only 34 (Fig. 2*D*). Third, *A. novofumigatus* also has more different types of clusters. An overview of the cluster types and the number of clusters found in the two species can be seen in Fig. 2*D*. The diversity of SM gene clusters supports the identification of these two organisms as separate species.

SMs can present a competitive advantage in the battle for resources, but if the environment is stable, there is no need for a large arsenal of different metabolites. Thus, the large difference in SM potential between *A. fumigatus* and *A. novofumigatus* might be a reflection of the difference in natural environment and the competition in these environments, indicating that *A. novofumigatus* normally exists in a highly competitive environment and has a need for a larger repertoire of SMs. These results do not suggest in which conditions the metabolites are produced. Perhaps the differences are influenced by that fact that *A. fumigatus* Af293 is from a clinical isolate, whereas *A. novofumigatus* have been isolated from chamise chaparral soil after a bush fire in Southern California (2, 24). Indeed, earlier analyses have shown that clinical isolates produce fewer exometabolites and sporulate less (4, 30, 31).

It is clear that the sequence of *A. novofumigatus* represents a significant number of unknown gene clusters, and thereby possibly interesting bioactive compounds. To start explore this treasure chest and to illustrate our approach of linking metabolites to their respective gene clusters, we have here identified four highly interesting compounds (novofumigatonin, *ent*-cycloechinulin, and *epi*-aszonalenin A and C) by liquid chromatography–mass spectrometry analysis (*SI Appendix*, Fig. S1), and we have identified the biosynthetic gene clusters by using comparative genomics. Our analysis targeted these four model compounds, as they represent major metabolites produced by *A. novofumigatus* and because we have them as pure standards in our in-house collection of fungal metabolites (32).

Novofumigatonin is chemically a very complex compound containing an orthoester and is at present only known to be produced by *A. novofumigatus* (33). It has been suggested that novofumigatonin

is a meroterpenoid produced from the aromatic polyketide, 3,5-dimethylorsellinic acid, as an initial precursor. Terretonin is another fungal meroterpenoid derived from 3,5-dimethylorsellinic acid, and thus we hypothesized that the early-stage biosynthetic route for novofumigatonin would follow the same pathway as that for terretonin, found in *A. terreus* (34).

The terretonin biosynthetic genes in *A. terreus* were used to find homologs in *A. novofumigatus*, using BLASTP. Exactly one candidate cluster was identified (Fig. 3*C*), containing orthologs to all of the genes from the early stages of the terretonin pathway, with one exception: a terpene cyclase (*trt1*). However, we found this just upstream of the identified predicted cluster, and it was thus included in the putative cluster and in Fig. 3*C*.

Novofumigatonin is closely related to fumigatonin, which has been reported to be produced by *A. fumigatus* (35). Interestingly, no similar cluster could be found in *A. fumigatus*. The most similar cluster in *A. fumigatus* (which is not a very strong hit: three proteins with amino acid identity <50%) is an already known cluster responsible for production of another meroterpenoid, pyripyropene A (36). This is very puzzling, and one could speculate that the report of fumigatonin might have been obtained from a misidentified isolate supposed to be *A. fumigatus* (35), but in reality an *A. novofumigatus* strain, which was only recently described as a separate species (24).

A putative cluster (scaffold 3, 33,642–53,133 bp) for *ent*-cycloechinulin (*SI Appendix*, Fig. S2) in *A. novofumigatus* was identified using the fumitremorgins (*ftm*) cluster from *A. fumigatus* as the starting point. Fumitremorgins are similar to *ent*-cycloechinulin (*SI Appendix*, Fig. S2), but with some important differences. *Ent*-cycloechinulin uses alanine instead of proline as a starter unit. Furthermore, the following prenylation occurs in a reverse manner. The genes responsible for these steps therefore have a low identity (≤35%). The following hydroxylation and *O*-methylation, however, are more similar, which is also reflected in the identity for the genes (≥65%). Even though the identity for the identified genes is low, they still represent the best hits in *A. novofumigatus*, supporting that this is the best candidate cluster.

In a similar fashion, a putative cluster for *epi*-aszonalenin A and C in *A. novofumigatus* was identified using a similar acetylaszonalenin cluster (*SI Appendix*, Fig. S2 for chemical structure) from *A. terreus*. Again, the acetylaszonalenin cluster proteins were used for a comparative genomics search in *A. novofumigatus*. This way, a very similar cluster was identified (scaffold 3, 1,663,448–1,719,848 bp) as a candidate for *epi*-aszonalenin A and C biosynthesis.

Comparative Genomics of Genes Encoding Allergens, Virulence, and Pathogenicity Factors. *A. fumigatus* is known to be a common opportunistic human pathogen (37), whereas *A. novofumigatus* has only been reported as pathogenic in one instance (5). This difference in known pathogenicity offers the opportunity to identify and compare allergens and genes involved in the pathogenicity based on orthology, and to gather insights into the potential harmfulness of *A. novofumigatus*.

A list of currently accepted allergenic proteins from the well-studied *A. fumigatus* can be extracted from the Allergome database (www.allergome.org). The sequences of the allergenic proteins from this list were compared with the annotated *A. novofumigatus* protein list using BLAST+, with parameters set to report full-length

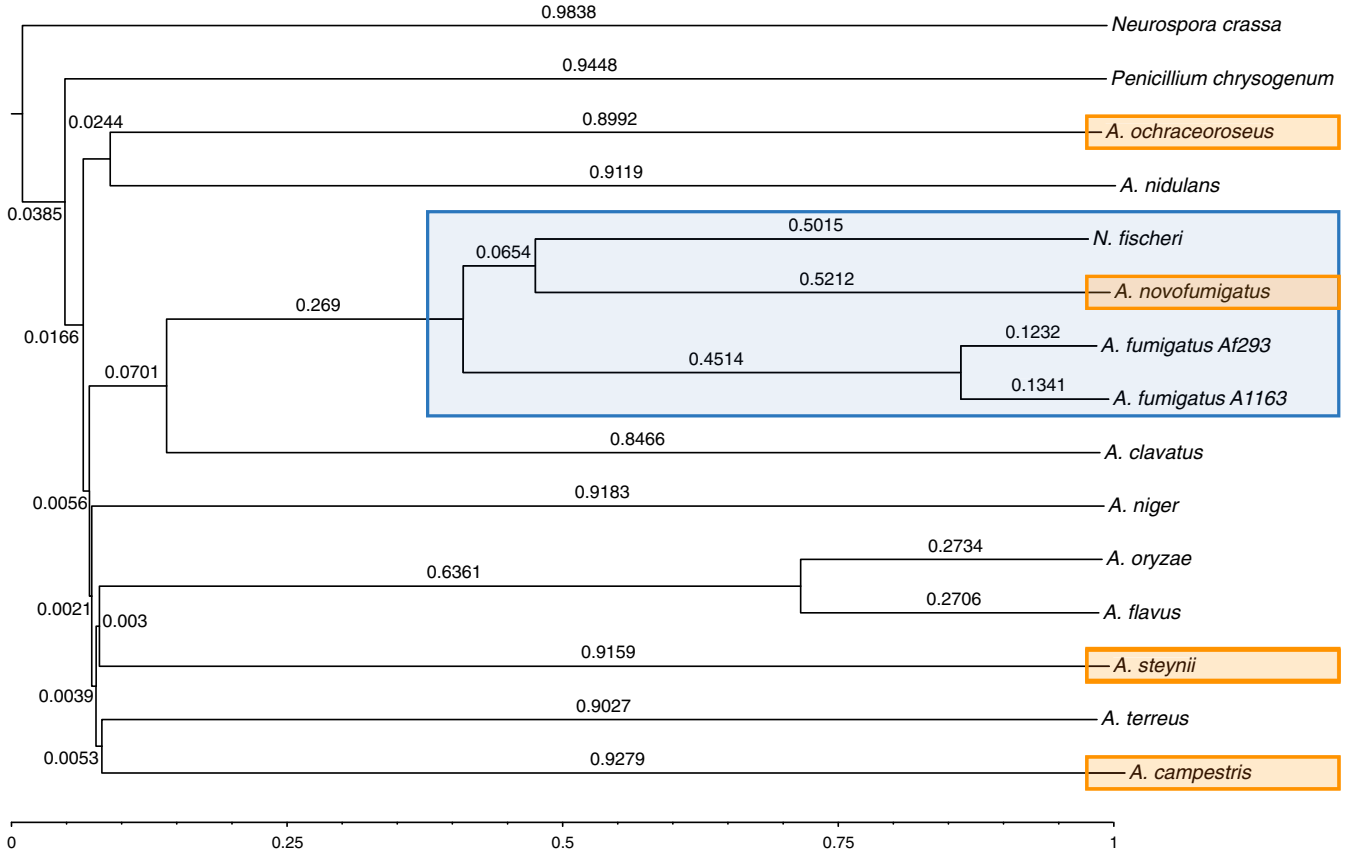


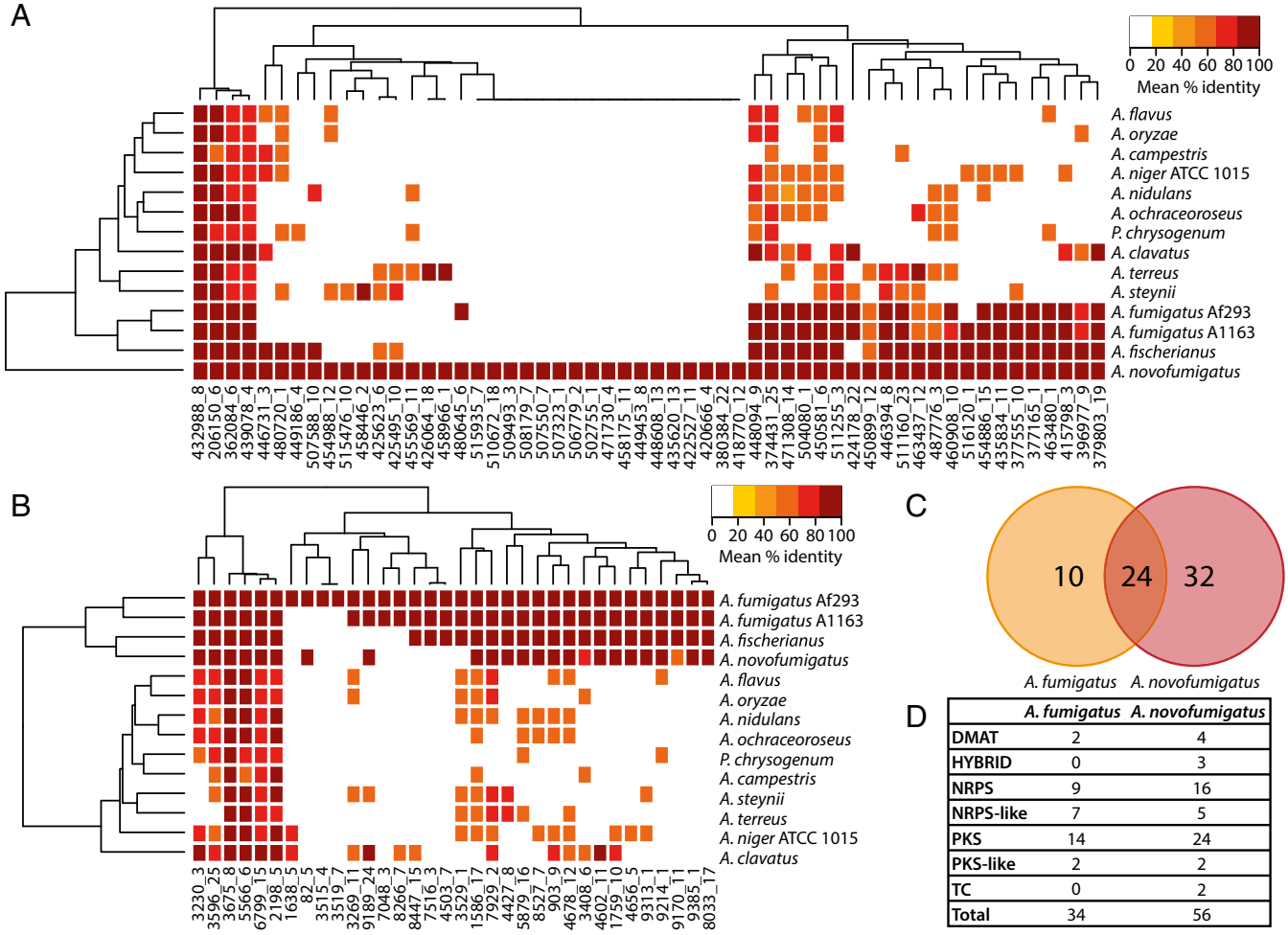
Fig. 1. Phylogenetic tree of 15 strains constructed from a composition vector approach of whole-proteome sequences using CVTree3 with a K'mer size of eight (19, 20). The time scale has been scaled to the root, thereby making the branch lengths relative to the distance between the root and the species. This time will therefore depend on what species are included in the set. The four PacBio-sequenced species are marked with orange, and the *Fumigati* section is marked with blue.

sequence matches. Results shown in *SI Appendix*, Table S2 indicate that all *A. fumigatus* allergen proteins are represented in the *A. novofumigatus* genome. Of a total of 41 proteins, 34 proteins showed >90% identity, four showed 85–90% identity, and three showed 50–80% identity. As proteins with >50% identity are likely to cross-react to IgE (38), these results strongly indicate that *A. novofumigatus* possesses a strong allergen repertoire that will at least cross-react strongly with IgE to *A. fumigatus* and is likely to be able to provoke an immune response in the same manner as *A. fumigatus*. It is not possible to rule out the possibility that *A. novofumigatus* could be a more virulent pathogen or allergenic sensitizer than *A. fumigatus*.

A set of 35 potential virulence genes was assembled from recent literature, as well as genes responsible for biosynthesis of the SMs melanin, fumagillin, fumitremorgins, gliotoxin, and helvolin, which are reported to play a direct role in virulence (4, 39, 40). The results are shown in *SI Appendix*, Table S3. The majority of the potential virulence genes are shared between *A. fumigatus* and *A. novofumigatus* with high similarity (>85% identity); only *arp2* and *gel2* had identity just below 50%. The fumitremorgins cluster consists of nine genes, six of which have identity <50%, including the synthase indicating that *A. novofumigatus* is unable to produce fumitremorgins. The two SM gene clusters for gliotoxin and fumagillin in *A. fumigatus* both have highly similar matches in *A. novofumigatus*. The cluster for helvolic acid has three genes of nine with low BLASTP identity of 42–48%. However, *A. novofumigatus* has been reported to produce helvolic acid, indicating that a high amino acid similarity of these genes is not required (4).

It is likely that different combinations of virulence factors among the species affect pathogenicity (31). It has been suggested that species unable to produce some metabolites may be able to produce proxy-exometabolites that can serve the same function. This could indicate that species producing many different kinds of exometabolites are potentially pathogenic (4).

A. novofumigatus possesses the full range of allergen proteins expressed by *A. fumigatus*, in addition to the majority of virulence factors including several SMs. Furthermore, *A. novofumigatus* has an extensive potential for SM production with 56 predicted gene clusters compared with 34 for *A. fumigatus*. Together, these results indicate that *A. novofumigatus* has a considerable potential to be pathogenic. The observation of only a single instance of invasive infection by *A. novofumigatus* (5) may result from the recent development of methods to identify this species, which has previously not been distinguishable from *A. fumigatus*. It has been found that ∼4–5% of *A. fumigatus* isolated from patients later turned out to be closely related species (41). Thus, the true pathogenic potential of *A. novofumigatus* might be underestimated. Similarly, allergen sensitization to *A. novofumigatus* is not currently



	A. fumigatus	*A. novofumigatus*
DMAT	2	4
HYBRID	0	3
NRPS	9	16
NRPS-like	7	5
PKS	14	24
PKS-like	2	2
TC	0	2
Total	34	56

Fig. 2. (*A*) Overview of the SM gene clusters predicted in *A. novofumigatus* and their homologs in the reference species. (*B*) Overview of the SM gene clusters predicted in *A. fumigatus* and their homologs in the reference species. (*C*) A Venn diagram of the *A. fumigatus* and *A. novofumigatus* SM gene clusters. (*D*) The number and different types of SM gene clusters predicted in *A. fumigatus* and *A. novofumigatus*. DMATS, dimethylallyl tryptophan synthase; NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase; TC, terpene cyclase.

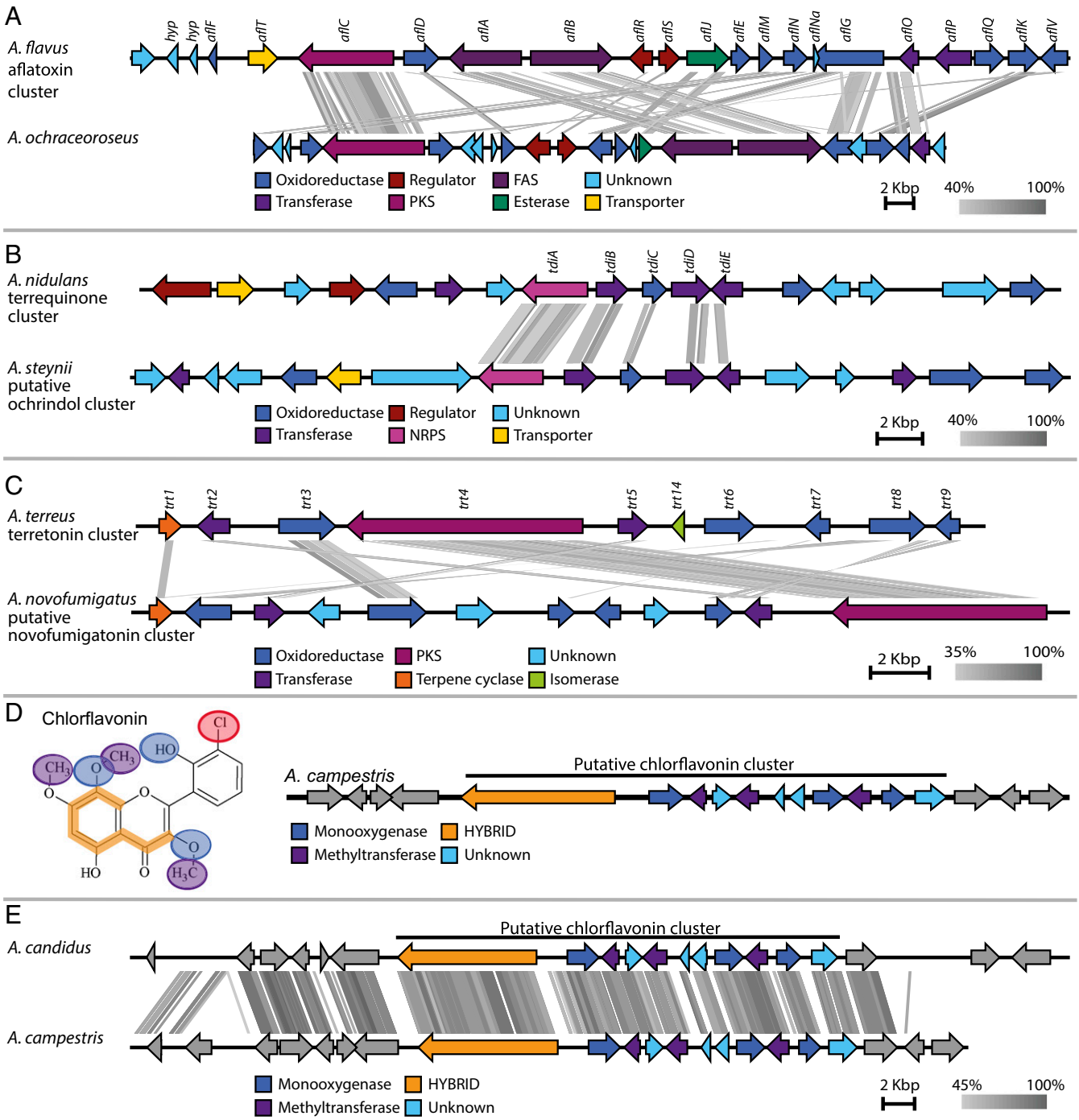


Fig. 3. Synteny plots of investigated clusters made using Easyfig tBLASTx. (*A*) The synteny of the predicted aflatoxin cluster in *A. flavus* NRRL3357 and the identified candidate aflatoxin cluster in *A. ochraceoroseus* (scaffold 2, 4,201,774–4,251,209 bp). (*B*) Synteny plot of the candidate cluster for ochrindol in *A. steynii* (scaffold 7, 2,783,445–2,824,507 bp) and terrequinone cluster in *A. nidulans*. The terrequinone cluster consists of a single-module nonribosomal peptide synthetase (tdiA), a prenyltransferase (tdiB), an oxidoreductase (tdiC), an aminotransferase (tdiD), and a gene of unknown function similar to a methyl transferase (tdiE). (*C*) Synteny plot of the known meroterpenoid cluster of terretonin in *A. terreus* and the candidate cluster of novofumigatonin in *A. novofumigatus* (scaffold 14, 103,246–136,450 bp). (*D*) The chemical structure of chlorflavonin and the candidate cluster for chlorflavonin in *A. campestris* (scaffold 1, 576,100–603,958 bp). The hydroxylation has been highlighted with blue, the O-methylation has been highlighted in purple, and the PKS backbone has been highlighted in orange. (*E*) Synteny plot of putative chlorflavonin clusters in *A. candidus* and *A. campestris*.

tested, and this species may also have potential to contribute to the burden of fungal allergy.

Investigation and Evolution of the Aflatoxin Gene Cluster in *A. ochraceoroseus*. It is well known that *A. ochraceoroseus* can produce aflatoxin, and the biosynthetic cluster has been identified (14). Also, it has been noted that the aflatoxin gene cluster in *A. ochraceoroseus* is missing homologs to the *aflP* and *aflQ* gene involved in the conversion of sterigmatocystin (ST) to aflatoxin.

Here we have compared the aflatoxin gene cluster from the whole-genome-sequenced *A. flavus* NRRL3357 with *A. ochraceoroseus*. The

clusters were identified in both species by using the aflatoxin genes identified in *A. flavus* AF70 (AY510453) (42).

Comparing the two clusters from *A. flavus* NRRL3357 with *A. ochraceoroseus*, it is evident that the synteny is characterized by gene shuffling (Fig. 3*A*). The identified cluster in *A. ochraceoroseus* is more similar to the ST gene cluster known from *A. nidulans* in the organization of genes, which was also the result of previous findings (14). This is evolutionary very interesting, as the clusters producing the same compound are quite different in their synteny, suggesting cluster dynamics or distant evolutionary origin.

As found by Cary et al. (14) it was seen that the *A. flavus aflP* and *aflQ* genes are missing in the *A. ochraceoroseus* aflatoxin cluster. These genes are important for the biosynthesis of aflatoxin. The whole-genome sequence was searched for orthologs to the *aflP* and *aflQ* genes from *A. flavus*, using BLASTP. The best hit for *aflQ* was JGI protein 547596, with identity 56.3% and coverage of 95.3%. The best hits for *aflP* were JGI proteins 430163, 506769, and 427152, with identity ranging from 40.5% to 36.6% and coverage between 31.4% and 37.7%. All the potential genes are located on a different scaffold than the aflatoxin cluster. The genes identified here are possible candidates for the *A. ochraceoroseus* version of the *aflP* and *aflQ* genes. With this information, it is not possible to determine exactly which genes are responsible for the conversion from ST to aflatoxin, but based on homology, these are the best candidates. Another possibility could be that the *aflP* and *aflQ* genes in *A. ochraceoroseus* have arisen via convergent evolution, and would thus not be found via homology analysis.

In summary, the identified aflatoxin gene cluster in this *A. ochraceoroseus* genome shows that *A. ochraceoroseus* and *A. flavus* most likely represent various stages of the aflatoxin cluster evolution. However, to get the full picture and truly understand the evolution of the clusters, more aflatoxin and sterigmatocystin producers need to be sequenced to be able to make bigger comparisons and get a better idea of where and when the different variations were created.

Identifying the Ochrindol Cluster in *A. steynii*. Ochrindoles are prenylated bisindolyl benzoid/quinone metabolites (*SI Appendix*, Fig. S2) that have shown anti-insectant properties (43), one reason that *A. steynii* is an interesting species. Ochrindol is produced by *A. steynii*, and the chemical structure is known, but the biosynthetic pathway is unknown (44). However, the biosynthesis of a similar compound, terrequinone (*SI Appendix*, Fig. S2), produced by *A. nidulans*, is known, and so are the five biosynthetic genes *tdiA–tdiE* (45). It has been shown that ochrindol D is produced as an intermediate during biosynthesis of terrequinone. We therefore hypothesize that the genes for the biosynthesis would be partly similar, and could thus be used to identify the ochrindol cluster.

First, the five *tdi* genes were identified in the *A. nidulans* genome in a predicted cluster consisting of 17 genes. Significantly, five genes similar to *A. nidulans tdiA–tdiE* were identified in a predicted cluster of 17 genes, with the synteny of the *tdiA–tdiE* orthologs conserved (Fig. 3*B* and *SI Appendix*, Table S4). However, none of the genes next to the five *tdi* genes showed any homology or synteny, suggesting that the size of the cluster is overpredicted, at least in *A. nidulans*. In *A. steynii*, some of the extra genes could be involved in ochrindol production.

Identifying the Chlorflavonin Cluster in *A. campestris*. Chlorflavonin was the first fully characterized flavone with fungal origin, and it is also the first naturally occurring flavone discovered to be chlorinated. It has been shown to have antifungal properties against specific species (46). The chemical structure of chlorflavonin (*SI Appendix*, Fig. S2) is known, and a biosynthetic pathway has been proposed, but no genes associated with the biosynthesis have been identified (47). With the whole-genome sequence for *A. campestris* at hand, we started exploring the genetic potential to identify the biosynthetic gene cluster responsible for producing this interesting compound.

Initially, looking at the chemical structure of this fungal flavonoid, an obvious idea for the biosynthesis would be that the backbone structure is created by a type III PKS, as the compound is so similar to plant flavonoids produced by type III PKS (48, 49). However, no type III PKS were found in *A. campestris*, suggesting a fungal-specific mode of biosynthesis. Next, investigating the chemical structure and proposed general biosynthesis for chlorflavonin (47), it could be deduced that the cluster must contain at least a PKS/hybrid backbone, three monooxygenases, three methyltransferases, and a chlorinating enzyme. Only one cluster met the requirements of three monooxygenases and three methyltransferases (Fig. 3*D*). The only concern with this candidate cluster is the lack of the essential chlorinating enzyme (*SI Appendix*, Table S5, Part 3).

First, sequences of known chlorinating enzymes (*SI Appendix*, Table S5, Part 1) were used to search for similar proteins in *A. campestris*, using BLASTP, but no genes were found (51–54). Second, relevant possible chlorinating InterPro domains were identified and found in four genes (*SI Appendix*, Table S5, Part 2), although it was not possible to pinpoint the best candidate of the chlorinating enzyme with these methods. However, the identified cluster is currently the best candidate cluster for chlorflavonin in *A. campestris*. Verification of this by knock-out experiments or heterologous expression could verify the candidate clusters as being responsible for the production of chlorflavonin, but this organism is not currently genetically engineerable, and the gene cluster is too large to transfer.

We therefore set out to support our prediction by sequencing and comparing genomes of several closely related species from section Candidi. *A. candidus* is a known chlorflavonin producer, whereas *A. taichungensis* is not (50). These species were therefore whole-genome sequenced to compare the pattern of the producers with the predicted clusters. *A. campestris*, *A. candidus*, and *A. taichungensis* each have 48, 45, and 43 predicted clusters. Based on the backbone, *A. campestris* and *A. candidus* share 35 clusters and *A. campestris* and *A. taichungensis* share 31 (BLASTP ≥50% identity and ≥130% hit+query coverage).

Comparing the genes found in the putative chlorflavonin cluster in *A. campestris* with the whole-genome sequences of *A. candidus* and *A. taichungensis*, *A. candidus* was homologous to genes in the putative chlorflavonin cluster (Fig. 3*E*). Moreover, this cluster is also the only cluster in *A. candidus* that has three methyltransferases and three monooxygenases. *A. taichungensis*, in contrast, does not have any significant hits of the predicted biosynthesis genes, as would be expected.

In addition, the chlorinating potential was investigated in these species. As with *A. campestris*, there were no BLASTP hits in *A. candidus* and *A. taichungensis* from the known chlorinating proteins (*SI Appendix*, Table S5, Part 1) (51–54).

Also, the possible chlorinating InterPro domains were investigated in the genomes of *A. candidus* and *A. taichungensis*. The number of hits were similar; however, *A. campestris* had one more hit for IPR001568, and both *A. campestirs* and *A. candidus* had one more hit for IPR008775, but none of the hits is found in SMURF-predicted clusters (*SI Appendix*, Table S5, Part 2).

These investigations further support that the identified cluster in *A. campestris* is the best candidate for chlorflavonin biosynthesis.

Conclusion

In this study, high-quality PacBio genome sequence data were generated for four *Aspergillus* species (*A. campestris*, *A. novofumigatus*, *A. ochraceoroseus*, and *A. steynii*) and investigated using comparative genomics. Furthermore, we have prepared draft genome sequences for two additional species: *A. taichungensis* and *A. candidus*.

These six species are diverse and represent various sections of the *Aspergillus* genus, and thereby provide insight into the genomic and biochemical diversity and potential of the genus.

The four PacBio sequenced species have been compared with a group of already whole-genome-sequenced *Aspergillus* species to determine the level of genetic diversity. A phylogram was constructed on the basis of the whole-genome proteomes, and the resulting tree supports the taxonomy of the genus and fits with a phylogenetic tree constructed by Peterson SW (21) and Kocsubé S, et al. (22), based on four loci or nine loci (21, 22). The tree confirms that *A. campestris*, *A. ochraceoroseus*, and *A. steynii* indeed represent sections of the *Aspergillus* genus, which have not been genome sequenced before. Analysis of the genomes show that these genomes represent a large number of species-specific genes, particularly within secondary metabolism.

Investigation of the presence of N6-methyldeoxyadenine of the four presented species shows very low levels of 6mA. Moreover, no 6mA sites were found symmetrically at ApTs, which has been found to be a characteristic feature of 6mA modification in early-diverging fungi (18), thus confirming previous suggestions that 6mA methylation is not significant in Aspergilli.

A. novofumigatus has been compared with a close relative, the pathogenic species *A. fumigatus*, to better understand the mechanism of pathogenicity and virulence. The predicted SM gene clusters were found to be very different for the two close relatives, with *A. novofumigatus* containing 65% more clusters than *A. fumigatus*.

All allergens known from *A. fumigatus* are also present in *A. novofumigatus*, and the majority of the virulence factors are shared between the two species. The major difference is that *A. novofumigatus* lacks the fumitremorgin cluster. However, it has been suggested that proxy-exometabolites may serve the same function and *A. novofumigatus* has an extensive arsenal of additional SM gene clusters. It is thus highly likely that *A. novofumigatus* is a highly capable pathogen.

Furthermore, we have, with multiple examples, demonstrated that it is possible to identify the respective gene cluster using whole-genome sequences if one has a well-established structure of a SM and biological and chemical insights to the pathway. This way we have reidentified the aflatoxin gene cluster in *A. ochraceoroseus*; the *epi*-aszonalenins, novofumigatonin, and *ent*-cycloechinulin gene clusters in *A. novofumigatus*; the ochrindol cluster in *A. steynii*; and finally, the chlorflavonin cluster in *A. campestris*, backed by additional info from sequencing the *A. taichungensis* and *A. candidus* genomes.

In summary, the six genome sequences presented in this study illustrate the large diversity found in the *Aspergillus* genus and highlight the potential for discovery of structurally diverse SMs. As our project of sequencing +300 species progresses along with other fungal genome sequencing projects (e.g., the 1K fungal genomes project 1000.fungalgenomes.org/home/), the potential for applying comparative genomics to get evolutionary insights and discover interesting SMs will only increase.

Materials and Methods

The materials include a list of sequenced strains. Methods include strain cultivation; genome sequencing, assembly, and annotation; DNA-methylation analysis; details for comparative genomics analysis; phylogeny; and chemical analysis of secondary metabolism. Details for all methods are found in *SI Appendix, SI Text*. In particular, we provide a detailed protocol for efficient, reproducible, and scalable DNA and RNA extraction from fungi.

ACKNOWLEDGMENTS. M.R.A. and T.C.V. gratefully acknowledge funding from the Villum Foundation, Grant VKR023437. Genome sequencing was kindly supported by Joint BioEnergy Institute and Joint Genome Institute. The work conducted by the US Department of Energy Joint Genome Institute, a US Department of Energy Office of Science User Facility, is supported by the Office of Science of the US Department of Energy under Contract DE-AC02-05CH11231. The US Department of Energy Joint BioEnergy Institute (www.jbei.org) is supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy.

1. Samson RA, et al. (2014) Phylogeny, identification and nomenclature of the genus Aspergillus. *Stud Mycol* 78:141–173.
2. Nierman WC, et al. (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. *Nature* 438:1151–1156.
3. Moore G, Mack B, Beltz S (2015) Draft genome sequences of two closely-related aflatoxigenic Aspergillus species obtained from the Ivory Coast. *Genome Biol Evol* 8:729–732.
4. Frisvad JC, Larsen TO (2016) Extrolites of Aspergillus fumigatus and other pathogenic species in Aspergillus section Fumigati. *Front Microbiol* 6:1485.
5. Peláez T, et al. (2013) Invasive aspergillosis caused by cryptic Aspergillus species: A report of two consecutive episodes in a patient with leukaemia. *J Med Microbiol* 62: 474–478.
6. Hoffmeister D, Keller NP (2007) Natural products of filamentous fungi: Enzymes, genes, and their regulation. *Nat Prod Rep* 24:393–416.
7. Inglis DO, et al. (2013) Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of Aspergillus nidulans, A. fumigatus, A. niger and A. oryzae. *BMC Microbiol* 13:91.
8. Frisvad JC, Larsen TO (2015) Chemodiversity in the genus Aspergillus. *Appl Microbiol Biotechnol* 99:7859–7877.
9. Perrin RM, et al. (2007) Transcriptional regulation of chemical diversity in Aspergillus fumigatus by LaeA. *PLoS Pathog* 3:e50.
10. Palmer JM, Keller NP (2010) Secondary metabolism in fungi: Does chromosomal location matter? *Curr Opin Microbiol* 13:431–436.
11. Brakhage AA, Schroeckh V (2011) Fungal secondary metabolites: Strategies to activate silent gene clusters. *Fungal Genet Biol* 48:15–22.
12. Osbourn A (2010) Secondary metabolic gene clusters: Evolutionary toolkits for chemical innovation. *Trends Genet* 26:449–457.
13. Yu J, et al. (2004) Clustered pathway genes in aflatoxin biosynthesis. *Appl Environ Microbiol* 70:1253–1262.
14. Cary JW, Ehrlich KC, Beltz SB, Harris-Coward P, Klich MA (2009) Characterization of the Aspergillus ochraceoroseus aflatoxin/sterigmatocystin biosynthetic gene cluster. *Mycologia* 101:352–362.
15. Grigoriev IV, Martinez DA, Salamov AA (2006) Fungal genomic annotation. *Appl Mycol Biotechnol* 6:123–142.
16. Machida M, et al. (2005) Genome sequencing and analysis of Aspergillus oryzae. *Nature* 438:1157–1161.
17. Wortman JR, et al. (2006) Whole genome comparison of the A. fumigatus family. *Med Mycol* 44:S3–S7.
18. Mondo SJ, et al. (2017) Widespread adenine N6-methylation of active genes in fungi. *Nat Genet* 49:964–968.
19. Qi J, Luo H, Hao B (2004) CVTree: A phylogenetic tree reconstruction tool based on whole genomes. *Nucleic Acids Res* 32:W45–W47.
20. Zuo G, Li Q, Hao B (2014) On K-peptide length in composition vector phylogeny of prokaryotes. *Comput Biol Chem* 53:166–173.
21. Peterson SW (2008) Phylogenetic analysis of Aspergillus species using DNA sequences from four loci. *Mycologia* 100:205–226.
22. Kocsubé S, et al. (2016) Aspergillus is monophyletic: Evidence from multiple gene phylogenies and extrolites profiles. *Stud Mycol* 85:199–213.
23. Mitchell A, et al. (2015) The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res* 43:D213–D221.
24. Hong S-B, Go S-J, Shin H-D, Frisvad JC, Samson RA (2005) Polyphasic taxonomy of Aspergillus fumigatus and related species. *Mycologia* 97:1316–1329.
25. Delcher AL, Phillippy A, Carlton J, Salzberg SL (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* 30:2478–2483.
26. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5:R12.
27. Delcher AL, et al. (1999) Alignment of whole genomes. *Nucleic Acids Res* 27: 2369–2376.
28. Khaldi N, et al. (2010) SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol* 47:736–741.
29. Medema MH, et al. (2015) Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* 11:625–631.
30. Frisvad JC, Samson RA (1990) Chemotaxonomy and morphology of Aspergillus fumigatus and related species. *Modern Concepts in Penicillium Apergillus Classification*, pp 201–208.
31. Tamiya H, et al. (2015) Secondary metabolite profiles and antifungal drug susceptibility of Aspergillus fumigatus and closely related species, Aspergillus lentulus, Aspergillus udagawae, and Aspergillus viridinutans. *J Infect Chemother* 21:385–391.
32. Nielsen KF, Månsson M, Rank C, Frisvad JC, Larsen TO (2011) Dereplication of microbial natural products by LC-DAD-TOFMS. *J Nat Prod* 74:2338–2348.
33. Rank C, et al. (2008) Novofumigatonin, a new orthoester meroterpenoid from Aspergillus novofumigatus. *Org Lett* 10:401–404.
34. Guo C-J, et al. (2012) Molecular genetic characterization of a cluster in A. terreus for biosynthesis of the meroterpenoid terretonin. *Org Lett* 14:5684–5687.
35. Okuyama E, Yamazaki M, Katsube Y (1984) Fumigatonin, a new meroterpenoid from Aspergillus fumigatus. *Tetrahedron Lett* 25:3233–3234.

36. Itoh T, et al. (2010) Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases. *Nat Chem* 2:858–864.
37. Latgé JP (1999) Aspergillus fumigatus and aspergillosis. *Clin Microbiol Rev* 12:310–350.
38. Aalberse RC, Akkerdaas J, van Ree R (2001) Cross-reactivity of IgE antibodies to allergens. *Allergy* 56:478–490.
39. Valiante V, Macheleidt J, Föge M, Brakhage AA (2015) The Aspergillus fumigatus cell wall integrity signaling pathway: Drug target, compensatory pathways, and virulence. *Front Microbiol* 6:325.
40. Rementeria A, et al. (2005) Genes and molecules involved in Aspergillus fumigatus virulence. *Rev Iberoam Micol* 22:1–23.
41. Hong S-B, et al. (2010) Re-identification of Aspergillus fumigatus sensu lato based on a new concept of species delimitation. *J Microbiol* 48:607–615.
42. Ehrlich KC, Yu J, Cotty PJ (2005) Aflatoxin biosynthesis gene clusters and flanking regions. *J Appl Microbiol* 99:518–527.
43. de Guzman FS, et al. (1994) Ochrindoles A-D: New bis-indolyl benzenoids from the sclerotia of Aspergillus ochraceus NRRL 3519. *J Nat Prod* 57:634–639.
44. Frisvad JC, Frank JM, Houbraken JAMP, Kuijpers AFA, Samson RA (2004) New ochratoxin A producing species of Aspergillus section Circumdati. *Stud Mycol* 50:23–43.
45. Balibar CJ, Howard-Jones AR, Walsh CT (2007) Terrequinone A biosynthesis through L-tryptophan oxidation, dimerization and bisprenylation. *Nat Chem Biol* 3:584–592.
46. Richards M, Bird AE, Munden JE (1969) Chlorflavonin, a new antifungal antibiotic. *J Antibiot* 22:388–389.
47. Burns MK, Coffin JM, Kurobane I, Vining LC (1979) Biosynthesis of chlorflavonin in Aspergillus candidus : A novel fungal route to flavonoids. *J Chem Soc Chem Commun* 426–427.
48. Hashimoto M, Nonaka T, Fujii I (2014) Fungal type III polyketide synthases. *Nat Prod Rep* 31:1306–1317.
49. Austin MB, Noel JP (2003) The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* 20:79–110.
50. Varga J, Frisvad JC, Samson RA (2007) Polyphasic taxonomy of Aspergillus section Candidi based on molecular, morphological and physiological data. *Stud Mycol* 59: 75–88.
51. Vaillancourt FH, Yeh E, Vosburg DA, O'Connor SE, Walsh CT (2005) Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosyn. *Nature* 436: 1191–1194.
52. Kirner S, et al. (1998) Functions encoded by pyrrolnitrin biosynthetic genes from Pseudomonas fluorescens. *J Bacteriol* 180:1939–1943.
53. Xu X, et al. (2014) Identification of the first diphenyl ether gene cluster for pestheic acid biosynthesis in plant endophyte Pestalotiopsis fici. *ChemBioChem* 15:284–292.
54. Fullone MR, et al. (2012) Insight into the structure-function relationship of the non-heme iron halogenases involved in the biosynthesis of 4-chlorothreonine–Thr3 from Streptomyces sp. OH-5093 and SyrB2 from Pseudomonas syringae pv. syringae B301DR. *FEBS J* 279:4269–4282.